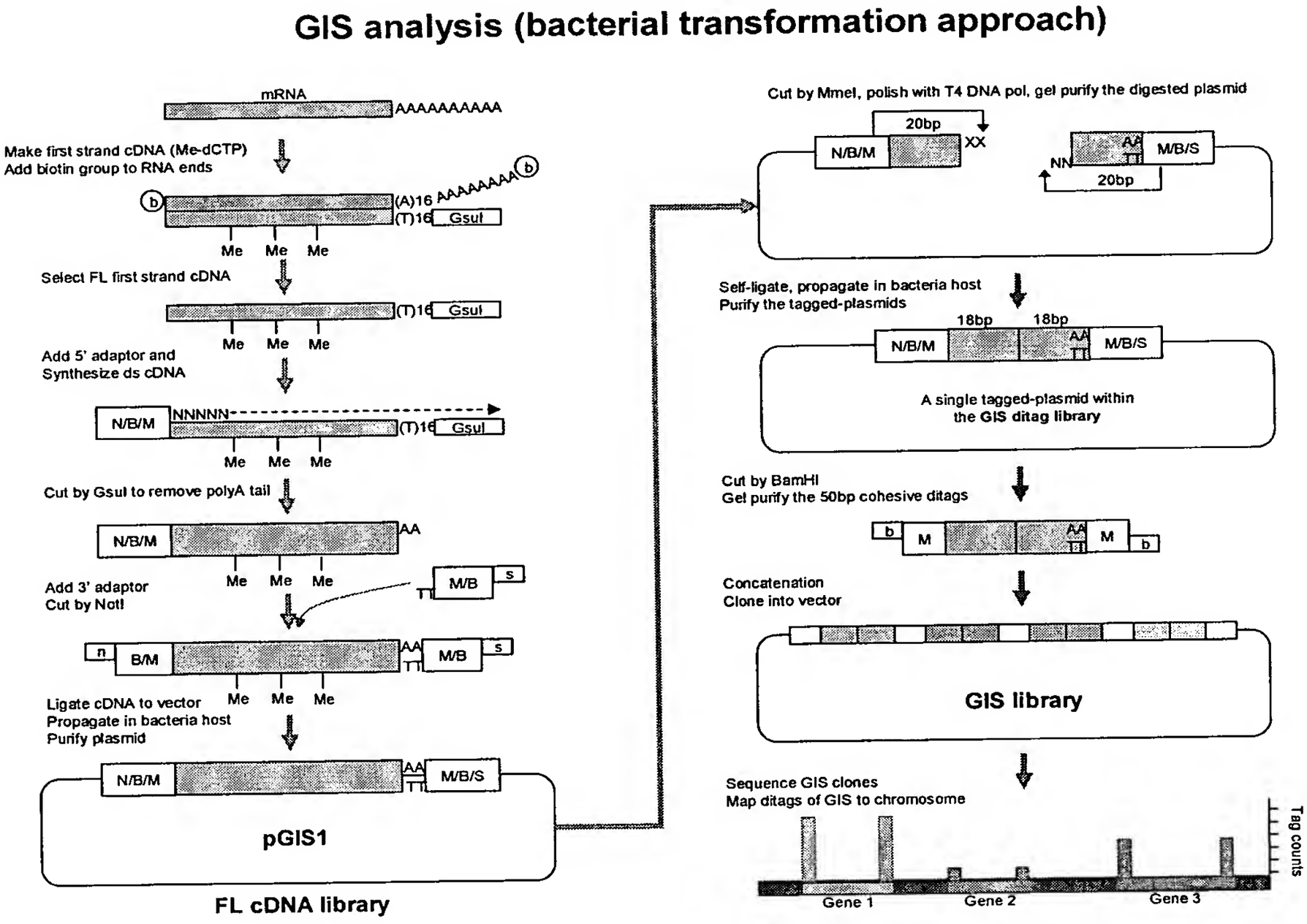
GIS analysis (bacterial transformation approach)
mRNA
AAAAAAAAAA
Make first strand cDNA (Me-dCTP)
Add biotin group to RNA ends
(A)16
(T)16
Gsul
Me
Select FL first strand cDNA
Add 5' adaptor and
Synthesize ds cDNA
N/B/M
NNNNN
Cut by Gsul to remove polyA tail
AA
Add 3' adaptor
Cut by NotI
M/B
S
n
B/M
Ligate cDNA to vector
Propagate in bacteria host
Purify plasmid
M/B/S
pGIS1
FL cDNA library
Cut by MmeI, polish with T4 DNA pol, gel purify the digested plasmid
20bp
XX
NN
Self-ligate, propagate in bacteria host
Purify the tagged-plasmids
18bp
A single tagged-plasmid within
the GIS ditag library
Cut by BamHI
Gel purify the 50bp cohesive ditags
b
M
Concatenation
Clone into vector
GIS library
Sequence GIS clones
Map ditags of GIS to chromosome
Tag counts
Gene 1
Gene 2
Gene 3

Figure 1

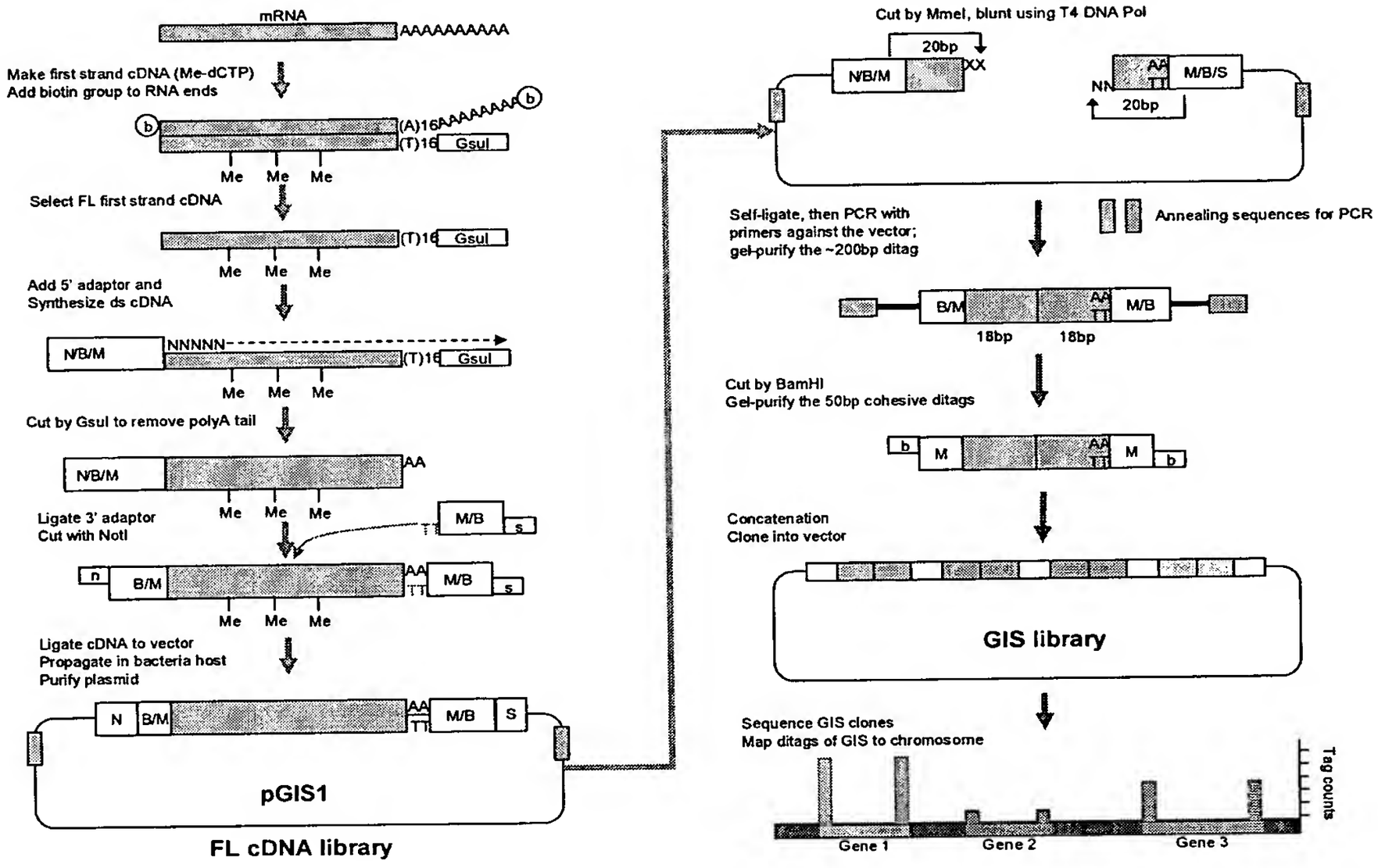
GIS analysis (PCR approach)
mRNA
AAAAAAAAAA
Make first strand cDNA (Me-dCTP)
Add biotin group to RNA ends
(A)16
(T)16
GsuI
Me
Select FL first strand cDNA
Add 5' adaptor and
Synthesize ds cDNA
N/B/M
NNNNN
Cut by GsuI to remove polyA tail
AA
Ligate 3' adaptor
Cut with NotI
M/B
S
B/M
TT
Ligate cDNA to vector
Propagate in bacteria host
Purify plasmid
N
pGIS1
FL cDNA library
Cut by MmeI, blunt using T4 DNA Pol
20bp
XX
NN
Self-ligate, then PCR with
primers against the vector;
gel-purify the ~200bp ditag
Annealing sequences for PCR
18bp
Cut by BamHI
Gel-purify the 50bp cohesive ditags
b
M
Concatenation
Clone into vector
GIS library
Sequence GIS clones
Map ditags of GIS to chromosome
Tag counts
Gene 1
Gene 2
Gene 3

Figure 2

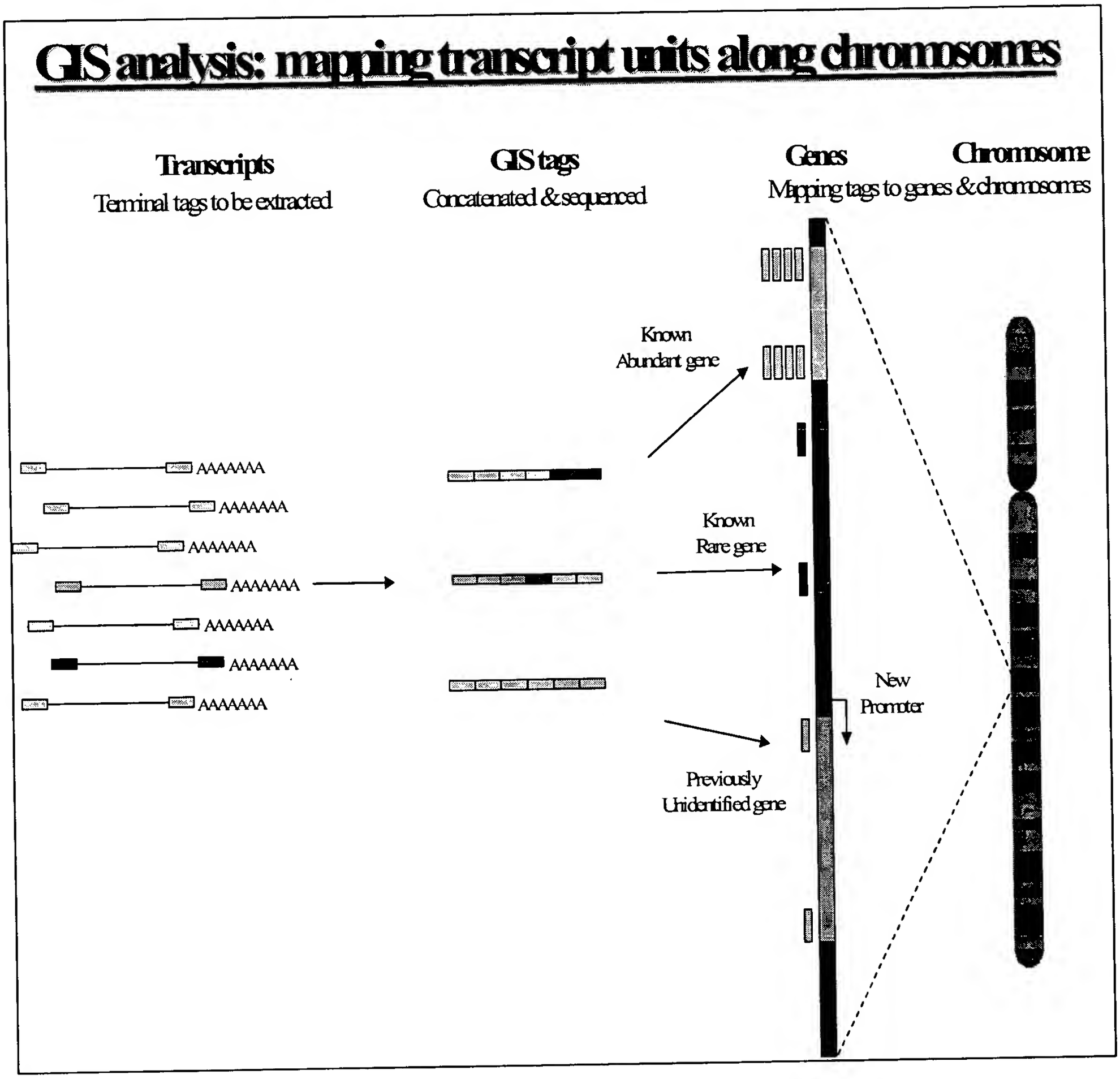
GIS analysis: mapping transcript units along chromosomes
Transcripts
Terminal tags to be extracted
GIS tags
Concatenated & sequenced
Genes
Chromsome
Mapping tags to genes & chromsomes
AAAAAAA
AAAAAAA
AAAAAAA
AAAAAAA
AAAAAAA
AAAAAAA
AAAAAAA
Known
Abundant gene
Known
Rare gene
New
Promoter
Previously
Unidentified gene

Figure 3

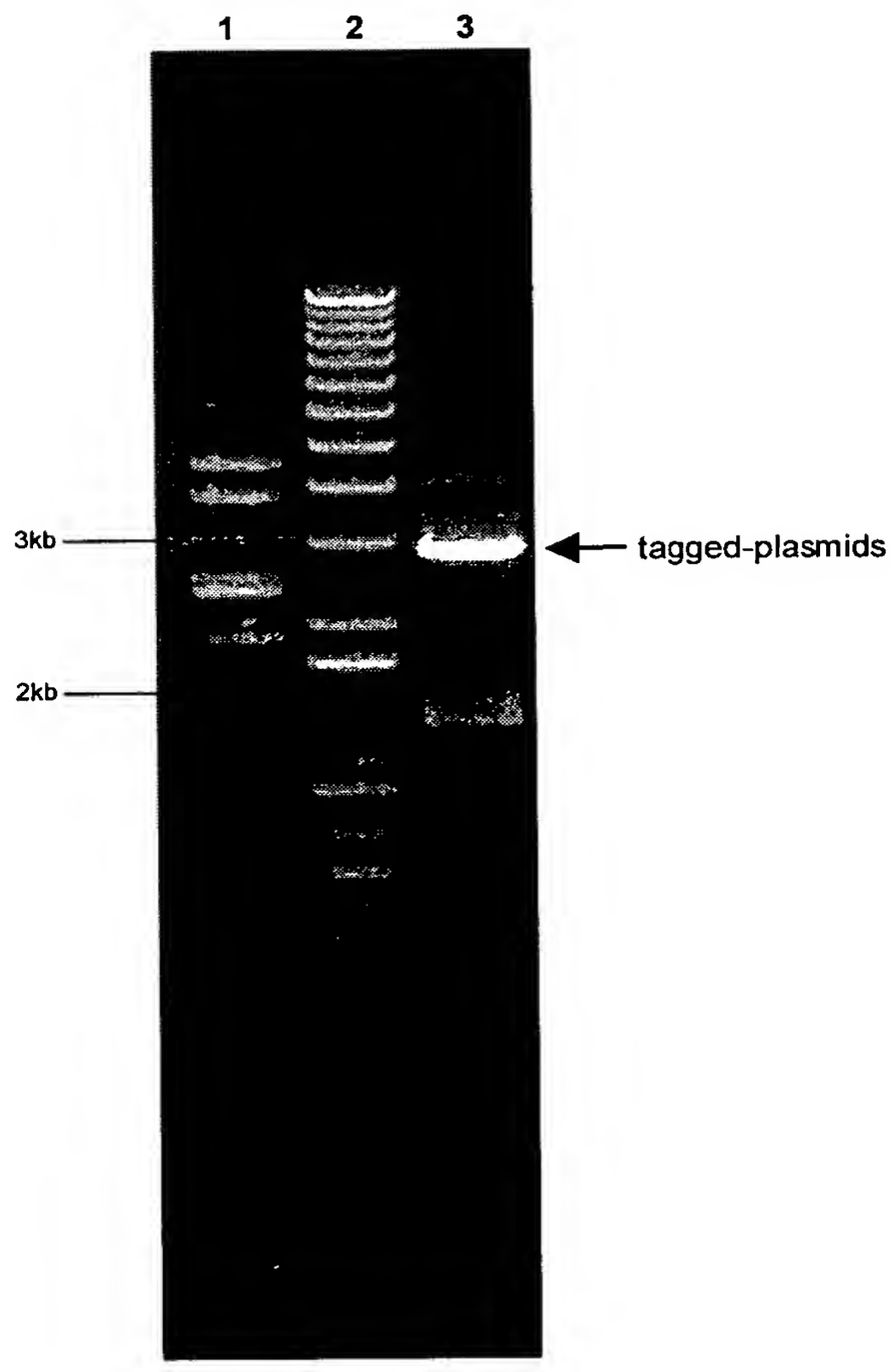

Figure 4

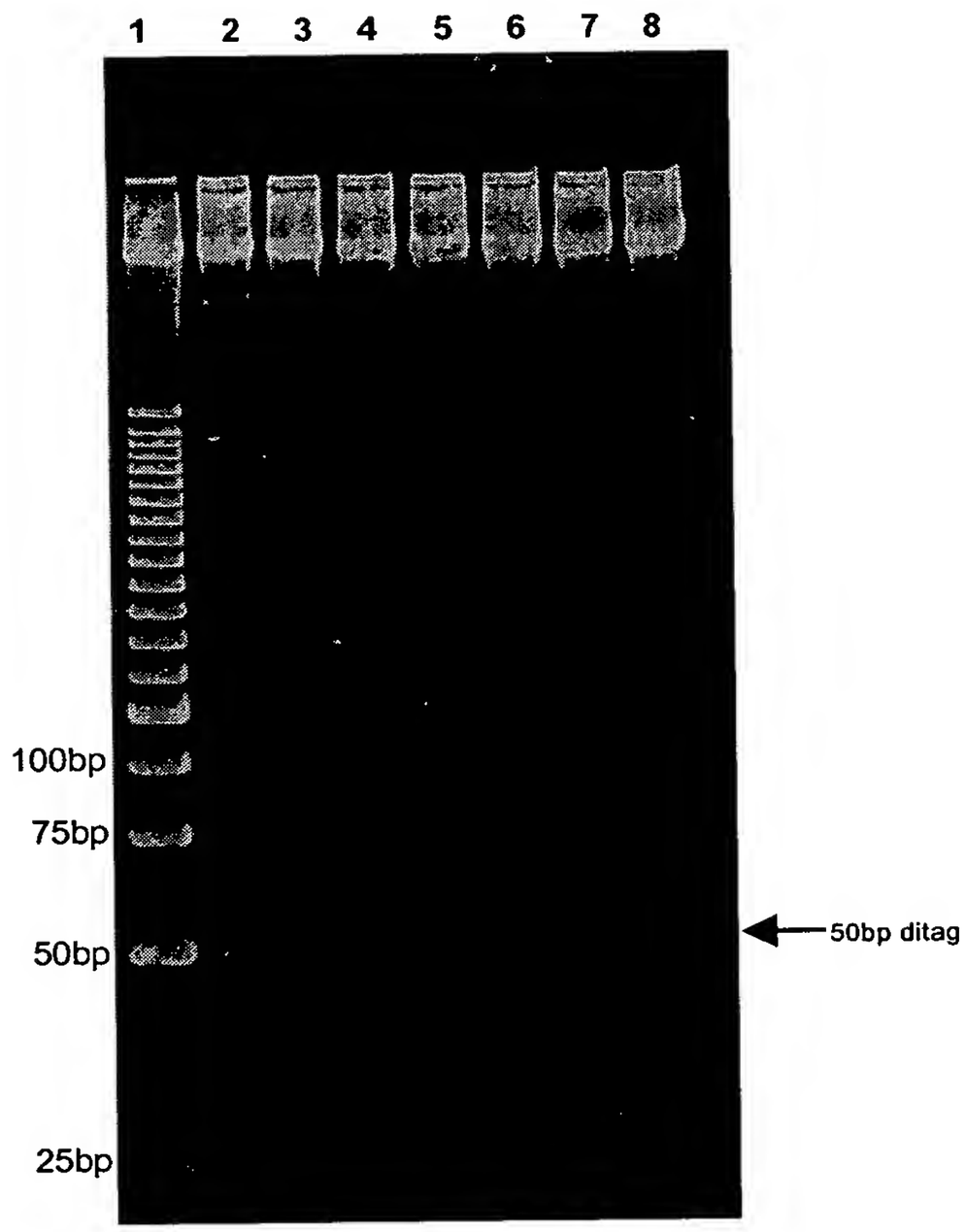

Figure 5

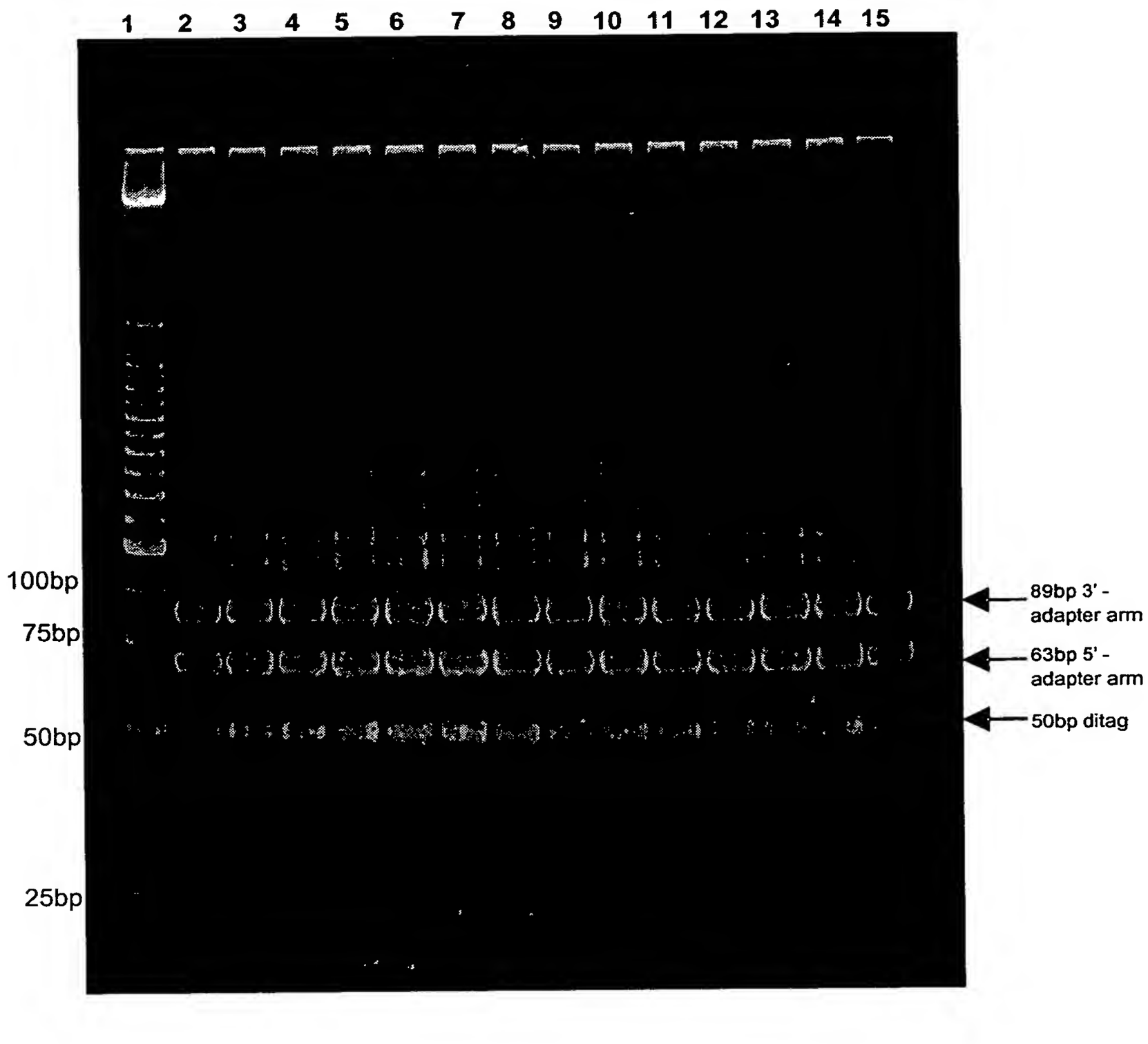
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
100bp
75bp
50bp
25bp
89bp 3' - adapter arm
63bp 5' - adapter arm
50bp ditag

Figure 6

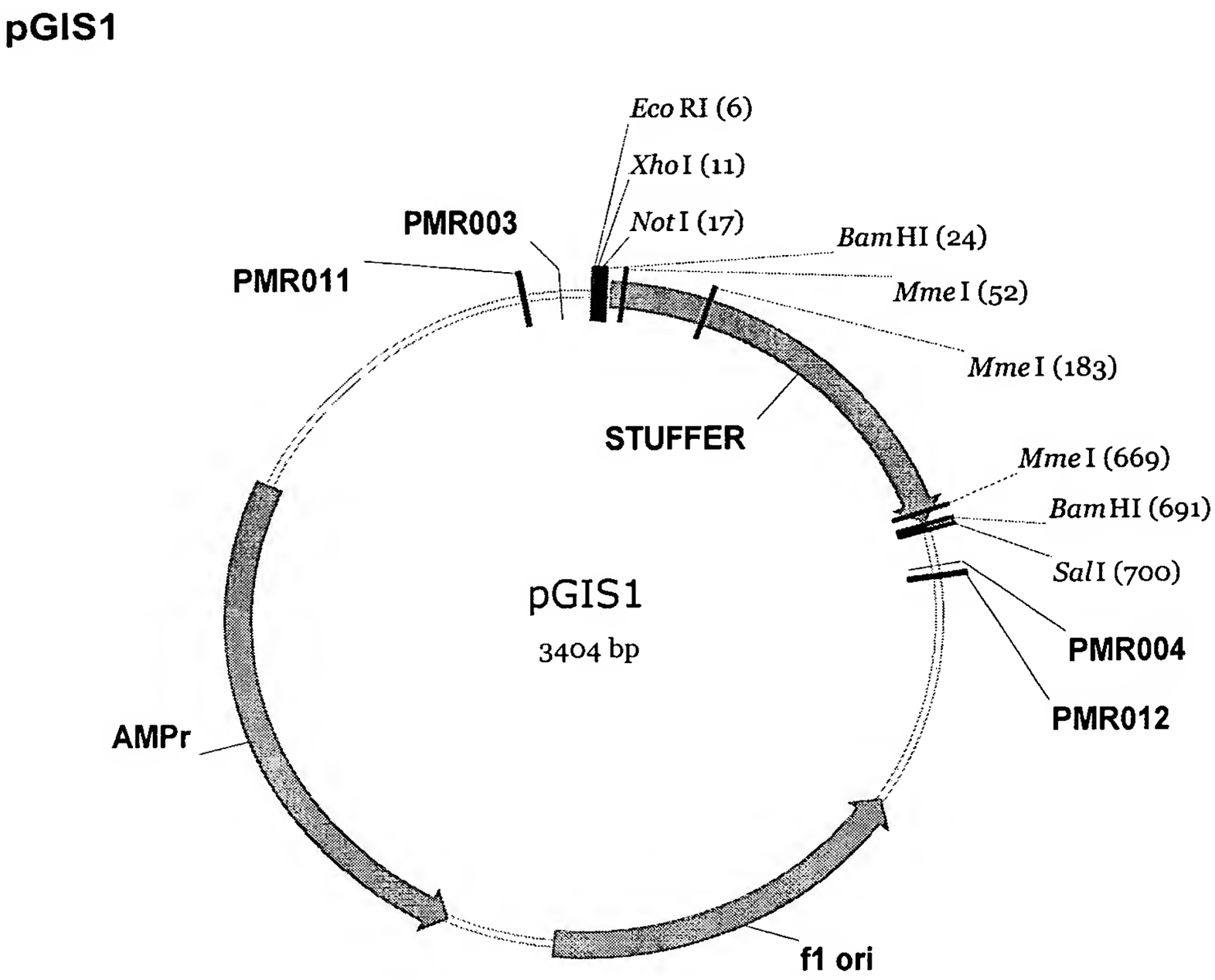
pGIS1
Eco RI (6)
Xho I (11)
Not I (17)
Bam HI (24)
Mme I (52)
Mme I (183)
Mme I (669)
Bam HI (691)
Sal I (700)
PMR003
PMR011
PMR004
PMR012
STUFFER
AMPr
f1 ori
pGIS1
3404 bp

Figure 7

pZErO-1

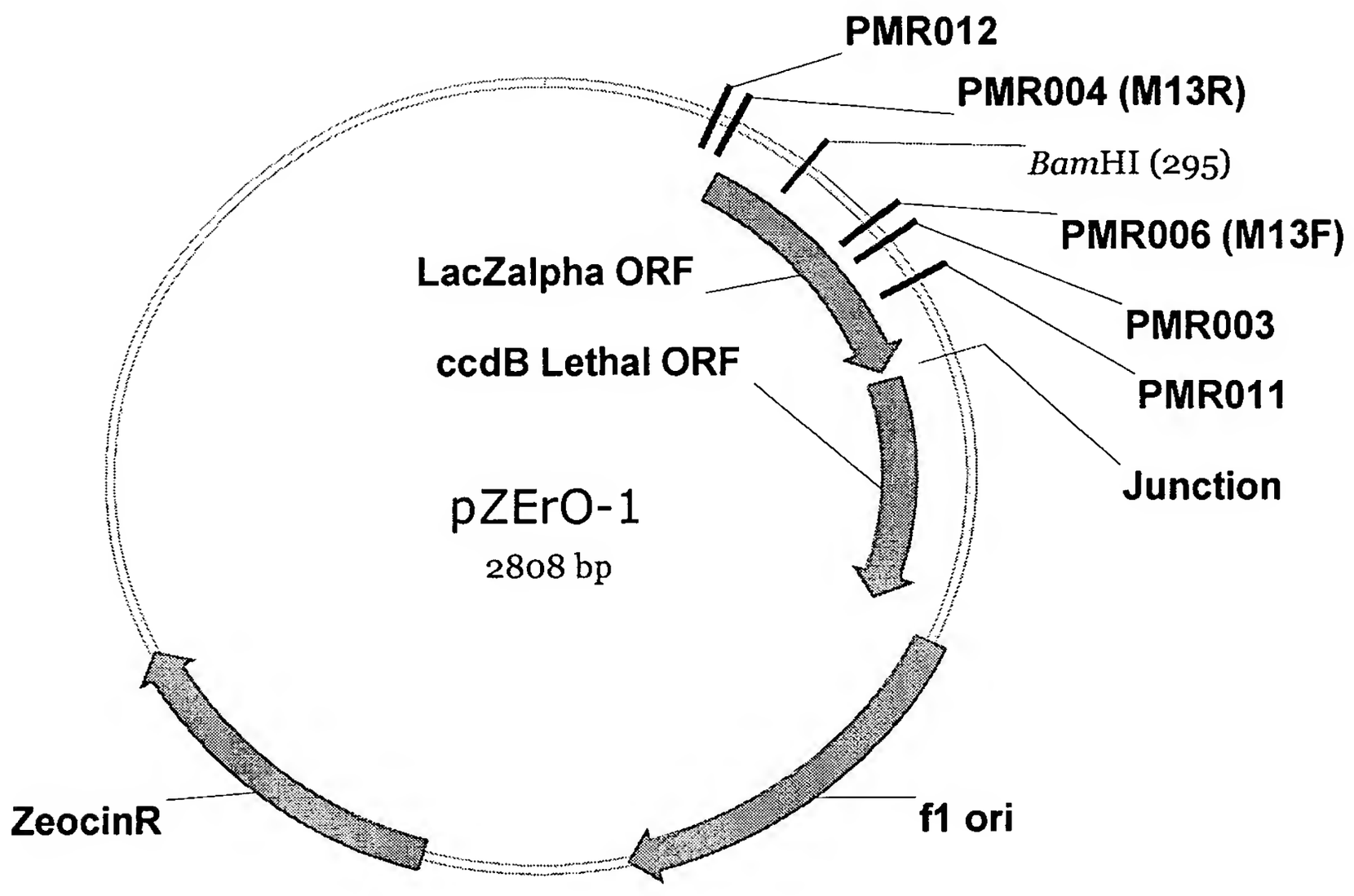

Figure 8

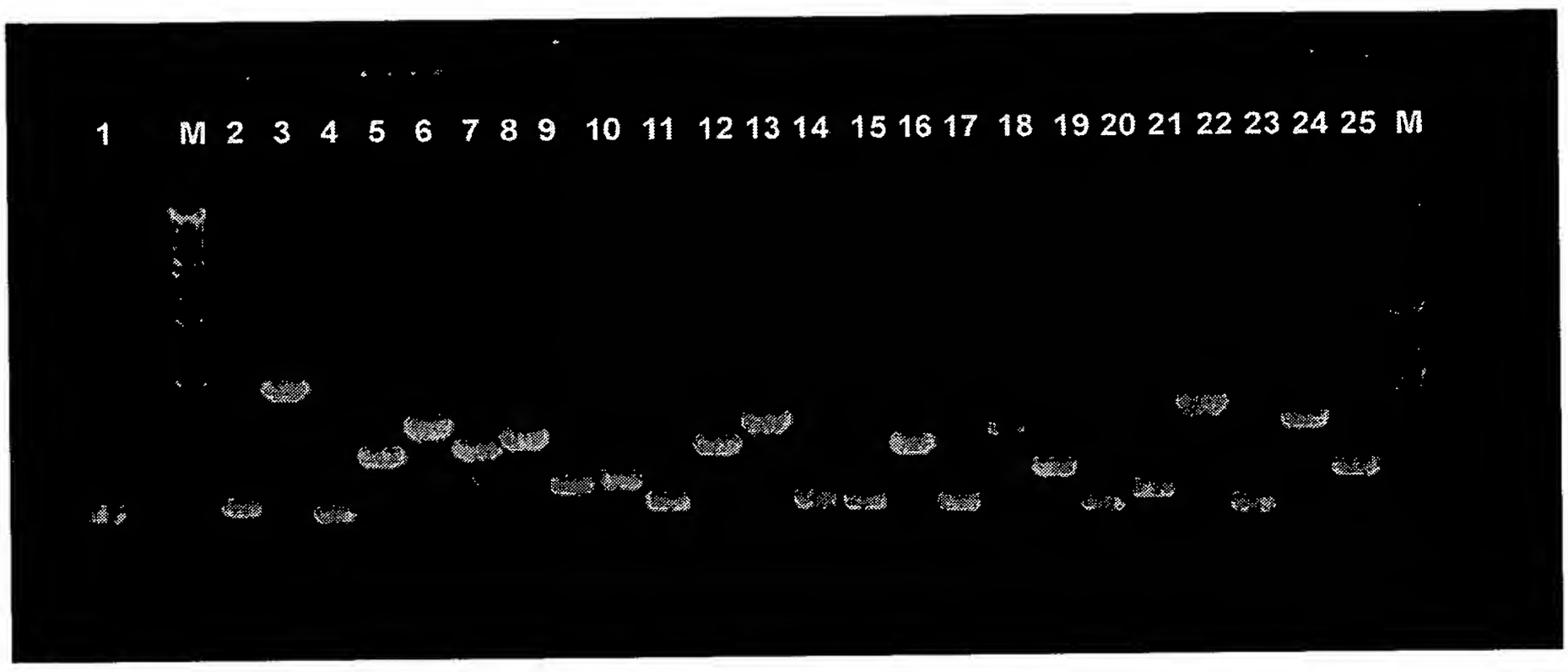

Figure 9

FIGURE 10
pGIS1 sequence

```
                                NotI
                              ---------
                      XhoI                MmeI
                    --------            -------
          EcoRI                   BamHI
        --------                  ------
1       GGGCGAATTC TCGAGCGGCC GCGGATCCGA CGAGAGCGCC TGCGTACGGC TCGCCGCGGT GGCTGGCGCT ACTTCGGAGG AGCCCGACGC GGCGCGGTCG
        CCCGCTTAAG AGCTCGCCGG CGCCTAGGCT GCTCTCGCGG ACGCATGCCG AGCGGCGCCA CCGACCGCGA TGAAGCCTCC TCGGGCTGCG CCGCGCCAGC

101     TTTTTATACA TTCCCGCGCG GAGGCAACGG AAGGGCGGGG CGCCTCGTGA TTAGGCCGCG GAGGTCACAG GCTCTGTTGT CATGAAGGTG AAAATTAAAT
        AAAAATATGT AAGGGCGCGC CTCCGTTGCC TTCCCGCCCC GCGGAGCACT AATCCGGCGC CTCCAGTGTC CGAGACAACA GTACTTCCAC TTTTAATTTA
          MmeI
        -------
201     GTTGGAATGG TGTGGCCACT TGGCTCTGGG TAGCCAATGA TGAGAACTGC GGCATCTGCA GGATGGCGTT TAATGGCTGC TGTCCAGACT GTAAGGTGCC
        CAACCTTACC ACACCGGTGA ACCGAGACCC ATCGGTTACT ACTCTTGACG CCGTAGACGT CCTACCGCAA ATTACCGACG ACAGGTCTGA CATTCCACGG

301     TGGTGATGAC TGCCCCCTCG TGTGGGGACA GTGCTCCCAC TGCTTCCACA TGCACTGCAT CCTCAAGTGG CTGAATGCGC AGCAGGTGCA GCAGCACTGC
        ACCACTACTG ACGGGGGAGC ACACCCCTGT CACGAGGGTG ACGAAGGTGT ACGTGACGTA GGAGTTCACC GACTTACGCG TCGTCCACGT CGTCGTGACG

401     CCCATGTGTC GCCAGGAGTG GAAGTTCAAA GAGTGAAGCC CGTGCCGTGC CACTTCCCTC TCCTGTGCTG TGCCAGGCTC AGCCCCTTCC CTCCCTCCCC
        GGGTACACAG CGGTCCTCAC CTTCAAGTTT CTCACTTCGG GCACGGCACG GTGAAGGGAG AGGACACGAC ACGGTCCGAG TCGGGGAAGG GAGGGAGGGG

501     TCCCCCAGAT ACAGCACCCC AAGTCCCCTC CACACAGCAC AGTGGTGCCC AGAGATCTCG GTCTGTGCCG GGGACAAGGA TGCTTTCTGT TTGGCTGGGA
        AGGGGGTCTA TGTCGTGGGG TTCAGGGGAG GTGTGTCGTG TCACCACGGG TCTCTAGAGC CAGACACGGC CCCTGTTCCT ACGAAAGACA AACCGACCCT
                                                                                                  MmeI
                                                                                                --------
                                                                                                    BamHI
                                                                                                  -------
601     CAAGGTTGAA AGGAGCTTTG CTGACTGTTT TGTTTTCCCA TCACATTGAC ACTTTATTCA ATAAGTAAAA CTCATTACAG TTCCAAGTCG GATCCTGGGT
        GTTCCAACTT TCCTCGAAAC GACTGACAAA ACAAAAGGGT AGTGTAACTG TGAAATAAGT TATTCATTTT GAGTAATGTC AAGGTTCAGC CTAGGACCCA
        SalI
        ----
701     CGACCTGCAG GCATGCAAGC TTGAGTATTC TATAGTGTCA CCTAAATAGC TTGGCGTAAT CATGGTCATA GCTGTTTCCT GTGTGAAATT GTTATCCGCT
        GCTGGACGTC CGTACGTTCG AACTCATAAG ATATCACAGT GGATTTATCG AACCGCATTA GTACCAGTAT CGACAAAGGA CACACTTTAA CAATAGGCGA

801     CACAATTCCA CACAACATAC GAGCCGGAAG CATAAAGTGT AAAGCCTGGG GTGCCTAATG AGTGAGCTAA CTCACATTAA TTGCGTTGCG CTCACTGCCC
        GTGTTAAGGT GTGTTGTATG CTCGGCCTTC GTATTTCACA TTTCGGACCC CACGGATTAC TCACTCGATT GAGTGTAATT AACGCAACGC GAGTGACGGG

901     GCTTTCCAGT CGGGAAACCT GTCGTGCCAG CTGCATTAAT GAATCGGCCA ACGCGCGGGG AGAGGCGGTT TGCGTATTGG GCGCTCTTCC GCTTCCTCGC
        CGAAAGGTCA GCCCTTTGGA CAGCACGGTC GACGTAATTA CTTAGCCGGT TGCGCGCCCC TCTCCGCCAA ACGCATAACC CGCGAGAAGG CGAAGGAGCG

1001    TCACTGACTC GCTGCGCTCG GTCGTTCGGC TGCGGCGAGC GGTATCAGCT CACTCAAAGG CGGTAATACG GTTATCCACA GAATCAGGGG ATAACGCAGG
        AGTGACTGAG CGACGCGAGC CAGCAAGCCG ACGCCGCTCG CCATAGTCGA GTGAGTTTCC GCCATTATGC CAATAGGTGT CTTAGTCCCC TATTGCGTCC

1101    AAAGAACATG TGAGCAAAAG GCCAGCAAAA GGCCAGGAAC CGTAAAAAGG CCGCGTTGCT GGCGTTTTTC GATAGGCTCC GCCCCCCTGA CGAGCATCAC
        TTTCTTGTAC ACTCGTTTTC CGGTCGTTTT CCGGTCCTTG GCATTTTTCC GGCGCAACGA CCGCAAAAAG CTATCCGAGG CGGGGGGACT GCTCGTAGTG

1201    AAAAATCGAC GCTCAAGTCA GAGGTGGCGA AACCCGACAG GACTATAAAG ATACCAGGCG TTTCCCCCTG GAAGCTCCCT CGTGCGCTCT CCTGTACCGA
        TTTTTAGCTG CGAGTTCAGT CTCCACCGCT TTGGGCTGTC CTGATATTTC TATGGTCCGC AAAGGGGGAC CTTCGAGGGA GCACGCGAGA GGACATGGCT

1301    CCCTGCCGCT TACCGGATAC CTGTCCGCCT TTCTCCCTTC GGGAAGCGTG GCGCTTTCTC ATAGCTCACG CTGTAGGTAT CTCAGTTCGG TGTAGGTCGT
        GGGACGGCGA ATGGCCTATG GACAGGCGGA AAGAGGGAAG CCCTTCGCAC CGCGAAAGAG TATCGAGTGC GACATCCATA GAGTCAAGCC ACATCCAGCA

1401    TCGCTCCAAG CTGGGCTGTG TGCACGAACC CCCCGTTCAG CCCGACCGCT GCGCCTTATC CGGTAACTAT CGTCTTGAGA CCAACCCGGT AAGACACGAC
        AGCGAGGTTC GACCCGACAC ACGTGCTTGG GGGGCAAGTC GGGCTGGCGA CGCGGAATAG GCCATTGATA GCAGAACTCT GGTTGGGCCA TTCTGTGCTG

1501    TTATCGCCAC TGGCAGCAGC CACTGGTAAC AGGATTAGCA GAGCGAGGTA TGTAGGCGGT GCTACAGAGT TCTTGAAGTG GTGGCCTAAC TACGGCTACA
        AATAGCGGTG ACCGTCGTCG GTGACCATTG TCCTAATCGT CTCGCTCCAT ACATCCGCCA CGATGTCTCA AGAACTTCAC CACCGGATTG ATGCCGATGT

1601    CTAGAAGGAC AGTATTTGGT ATCTGCGCTC TGCTGAAGCC AGTTACCTTC GGAAAAAGAG TTGGTAGCTC TTGATCCGGC AAACAAACCA CCGCTGGTAG
        GATCTTCCTG TCATAAACCA TAGACGCGAG ACGACTTCGG TCAATGGAAG CCTTTTTCTC AACCATCGAG AACTAGGCCG TTTGTTTGGT GGCGACCATC

1701    CGGTGGTTTT TTTGTTTGCA AGCAGCAGAT TACGCGCAGA AAAAAAGGAT CTCAAGAAGA TCCTTTGATC TTTTCTACGG GGTCTGACGC TCAGTGGAAC
        GCCACCAAAA AAACAAACGT TCGTCGTCTA ATGCGCGTCT TTTTTTCCTA GAGTTCTTCT AGGAAACTAG AAAAGATGCC CCAGACTGCG AGTCACCTTG

1801    GAAAACTCAC GTTAAGGGAT TTTGGTCATG AGATTATCAA AAAGGATCTT CACCTAGATC CTTTTAAATT AAAAATGAAG TTTTAAATCA ATCTAAAGTA
        CTTTTGAGTG CAATTCCCTA AAACCAGTAC TCTAATAGTT TTTCCTAGAA GTGGATCTAG GAAAATTTAA TTTTTACTTC AAAATTTAGT TAGATTTCAT

1901    TATATGAGTA AACTTGGTCT GACAGTTACC AATGCTTAAT CAGTGAGGCA CCTATCTCAG CGATCTGTCT ATTTCGTTCA TCCATAGTTG CCTGACTCCC
        ATATACTCAT TTGAACCAGA CTGTCAATGG TTACGAATTA GTCACTCCGT GGATAGAGTC GCTAGACAGA TAAAGCAAGT AGGTATCAAC GGACTGAGGG

2001    CGTCGTGTAG ATAACTACGA TACGGGAGGG CTTACCATCT GGCCCCAGTG CTGCAATGAT ACCGCGAGAC CCACGCTCAC CGGCTCCAGA TTTATCAGCA
        GCAGCACATC TATTGATGCT ATGCCCTCCC GAATGGTAGA CCGGGGTCAC GACGTTACTA TGGCGCTCTG GGTGCGAGTG GCCGAGGTCT AAATAGTCGT

2101    ATAAACCAGC CAGCCGGAAG GGCCGAGCGC AGAAGTGGTC CTGCAACTTT ATCCGCCTCC ATCCAGTCTA TTAATTGTTG CCGGGAAGCT AGAGTAAGTA
        TATTTGGTCG GTCGGCCTTC CCGGCTCGCG TCTTCACCAG GACGTTGAAA TAGGCGGAGG TAGGTCAGAT AATTAACAAC GGCCCTTCGA TCTCATTCAT

2201    GTTCGCCAGT TAATAGTTTG CGCAACGTTG TTGGCATTGC TACAGGCATC GTGGTGTCAC GCTCGTCGTT TGGTATGGCT TCATTCAGCT CCGGTTCCCA
        CAAGCGGTCA ATTATCAAAC GCGTTGCAAC AACCGTAACG ATGTCCGTAG CACCACAGTG CGAGCAGCAA ACCATACCGA AGTAAGTCGA GGCCAAGGGT

2301    ACGATCAAGG CGAGTTACAT GATCCCCCAT GTTGTGCAAA AAAGCGGTTA GCTCCTTCGG TCCTCCGATC GTTGTCAGAA GTAAGTTGGC CGCAGTGTTA
        TGCTAGTTCC GCTCAATGTA CTAGGGGGTA CAACACGTTT TTTCGCCAAT CGAGGAAGCC AGGAGGCTAG CAACAGTCTT CATTCAACCG GCGTCACAAT

2401    TCACTCATGG TTATGGCAGC ACTGCATAAT TCTCTTACTG TCATGCCATC CGTAAGATGC TTTTCTGTGA CTGGTGAGTA CTCAACCAAG TCATTCTGAG
        AGTGAGTACC AATACCGTCG TGACGTATTA AGAGAATGAC AGTACGGTAG GCATTCTACG AAAAGACACT GACCACTCAT GAGTTGGTTC AGTAAGACTC

2501    AATAGTGTAT GCGGCGACCG AGTTGCTCTT GCCCGGCGTC AATACGGGAT AATACCGCGC CACATAGCAG AACTTTAAAA GTGCTCATCA TTGGAAAACG
        TTATCACATA CGCCGCTGGC TCAACGAGAA CGGGCCGCAG TTATGCCCTA TTATGGCGCG GTGTATCGTC TTGAAATTTT CACGAGTAGT AACCTTTTGC

2601    TTCTTCGGGG CGAAAACTCT CAAGGATCTT ACCGCTGTTG AGATCCAGTT CGATGTAACC CACTCGTGCA CCCAACTGAT CTTCAGCATC TTTTACTTTC
        AAGAAGCCCC GCTTTTGAGA GTTCCTAGAA TGGCGACAAC TCTAGGTCAA GCTACATTGG GTGAGCACGT GGGTTGACTA GAAGTCGTAG AAAATGAAAG

2701    ACCAGCGTTT CTGGGTGAGC AAAAACAGGA AGGCAAAATG CCGCAAAAAA GGGAATAAGG GCGACACGGA AATGTTGAAT ACTCATACTC TTCCTTTTTC
        TGGTCGCAAA GACCCACTCG TTTTTGTCCT TCCGTTTTAC GGCGTTTTTT CCCTTATTCC CGCTGTGCCT TTACAACTTA TGAGTATGAG AAGGAAAAAG

2801    AATATTATTG AAGCATTTAT CAGGGTTATT GTCTCATGAG CGGATACATA TTTGAATGTA TTTAGAAAAA TAAACAAATA GGGGTTCCGC GCACATTTCC
        TTATAATAAC TTCGTAAATA GTCCCAATAA CAGAGTACTC GCCTATGTAT AAACTTACAT AAATCTTTTT ATTTGTTTAT CCCCAAGGCG CGTGTAAAGG

2901    CCGAAAAGTG CCACCTGACG TCTAAGAAAC CATTATTATC ATGACATTAA CCTATAAAAA TAGGCGTATC ACGAGGCCCT TTCGTCTCGC GCGTTTCGGT
        GGCTTTTCAC GGTGGACTGC AGATTCTTTG GTAATAATAG TACTGTAATT GGATATTTTT ATCCGCATAG TGCTCCGGGA AAGCAGAGCG CGCAAAGCCA

3001    GATGACGGTG AAAACCTCTG ACACATGCAG CTCCCGGAGA CGGTCACAGC TTGTCTGTAA GCGGATGCCG GGAGCAGACA AGCCCGTCAG GGCGCGTCAG
        CTACTGCCAC TTTTGGAGAC TGTGTACGTC GAGGGCCTCT GCCAGTGTCG AACAGACATT CGCCTACGGC CCTCGTCTGT TCGGGCAGTC CCGCGCAGTC

3101    CGGGTGTTGG CGGGTGTCGG GGCTGGCTTA ACTATGCGGC ATCAGAGCAG ATTGTACTGA GAGTGCACCA TATGCGGTGT GAAATACCGC ACAGATGCGT
        GCCCACAACC GCCCACAGCC CCGACCGAAT TGATACGCCG TAGTCTCGTC TAACATGACT CTCACGTGGT ATACGCCACA CTTTATGGCG TGTCTACGCA

3201    AAGGAGAAAA TACCGCATCA GGCGCCATTC GCCATTCAGG CTGCGCAACT GTTGGGAAGG GCGATCGGTG CGGGCCTCTT CGCTATTACG CCAGCTGGCG
        TTCCTCTTTT ATGGCGTAGT CCGCGGTAAG CGGTAAGTCC GACGCGTTGA CAACCCTTCC CGCTAGCCAC GCCCGGAGAA GCGATAATGC GGTCGACCGC

3301    AAAGGGGGAT GTGCTGCAAG GCGATTAAGT TGGGTAACGC CAGGGTTTTC CCAGTCACGA CGTTGTAAAA CGACGGCCAG TGAATTGTAA TACGACTCAC
        TTTCCCCCTA CACGACGTTC CGCTAATTCA ACCCATTGCG GTCCCAAAAG GGTCAGTGCT GCAACATTTT GCTGCCGGTC ACTTAACATT ATGCTGAGTG

3401    TATA
        ATAT
```